Attention and Talker-Specificity in the Memory Encoding of Spoken Sentences

William Clapp
Meghan Sumner
Stanford University

LSA
Philadelphia
January 10, 2025

National Science Foundation
WOD FOUNDATION
Josephine de Karman Fellowship

Background

Talker-specific, ***acoustically-detailed*** memory for individual words.

(Bradlow et al., 1999; Goldinger, 1996; Palmeri et al., 1993)

- Better memory for ***same*** talker than ***different*** talker.
- Speech rate, intonation, emotion, extrinsic noise. . .

  (Bradlow et al., 1999; Nygaard & Queen, 2008; Pufahl & Samuel, 2014; Sheffert, 1998)

Memory is central to language understanding.

(Goldinger, 1998; Pierrehumbert, 2016; Wedel, 2012)

Problem

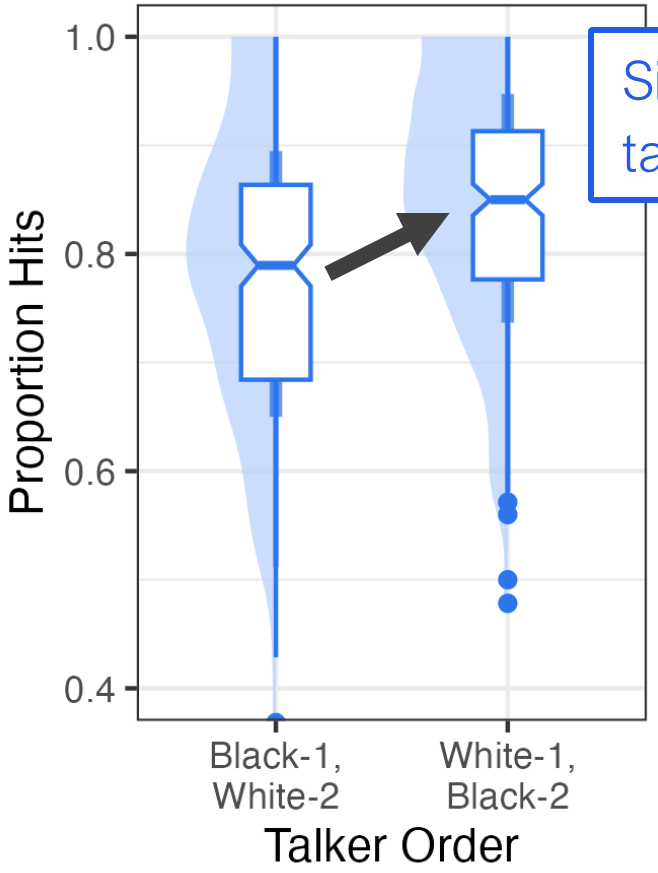
Most speech experiences are ***more complicated!***

- Longer utterances.
- Multi-tasking; planning responses.
- Talker information and messages interact in complex ways.

Fine-grained info is critical at the ***word level.***

How explanatory is this in ***longer utterance*** with ***various cognitive demands?***

Problem



Clapp, Vaughn, & Sumner, 2023

Simply swapping the order of talkers, memory patterns change.

***Memory Asymmetries**:* Some utterances are remembered better than others.

Asymmetries may result from ***resource allocation*** related to cognitive demands.

If specificity and memory asymmetries are ***fundamental*** to the system, they should be evident ***beyond the word.***

Question

How does dynamic resource allocation shape *linguistic representations?*

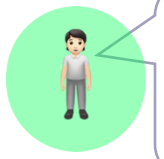
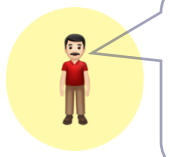
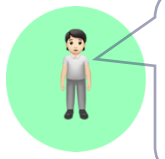
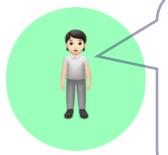
Is *talker-specific* information stored in memory for *full sentences*? (Exp. 1 – Validation)

What is the effect of *resource allocation* on memory for *talker-specific* information? (Exp. 1)

What is the effect of resource allocation on memory for sentences' *linguistic/conceptual* information? (Exps. 2 & 3)

Current study

Recognition Memory with *Full* or *Divided* Attention:

		Study		*Test*
Exp. 1	Audio sentences	 *"The front desk is attended by an elderly man."*	→ Audio	 *"The front desk is attended by an elderly man."*
Exp. 2	Audio sentences	 *"The front desk is attended by an elderly man."*	→ Text	The front desk is attended by an elderly man.
Exp. 3	Audio sentences	 *"The front desk is attended by an elderly man."*	→ Images	

How does resource allocation affect talker-specific memory for sentences?

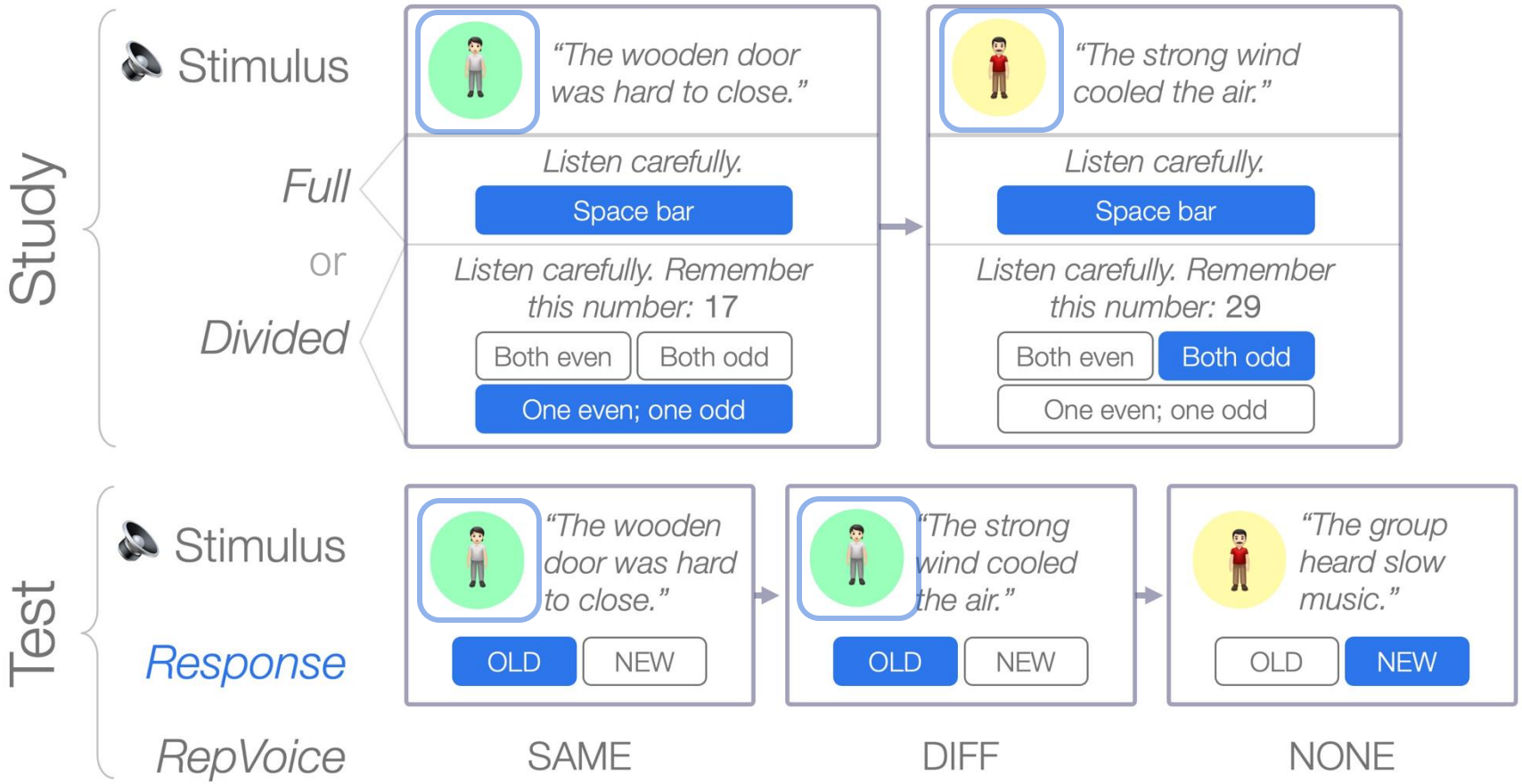
Design – Audio

Participants: From Prolific; Full (*N* = 163), Divided (*N* = 159).

Talkers: *2 female; 2 male GA speakers.*

Stimuli: *Basic English Lexicon sentence list (Rimikis, Smiljanic, & Calandruccio, 2013)*

RepVoice: *SAME vs. DIFF talker.*



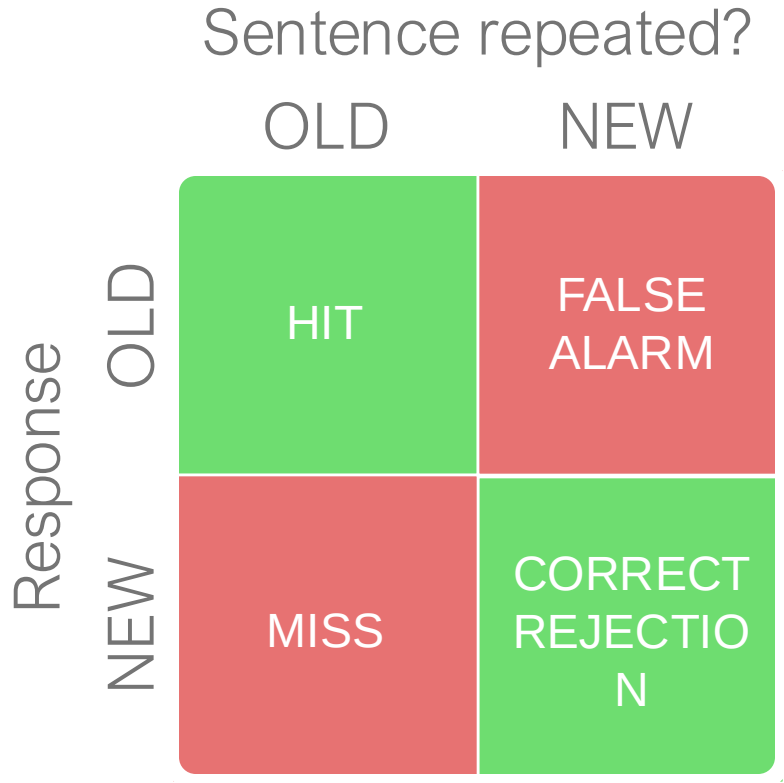
Analysis

Hits: OLD responses on OLD sentences.

False alarms: OLD responses on NEW sentences.

D': z(hits) – z(false alarms)

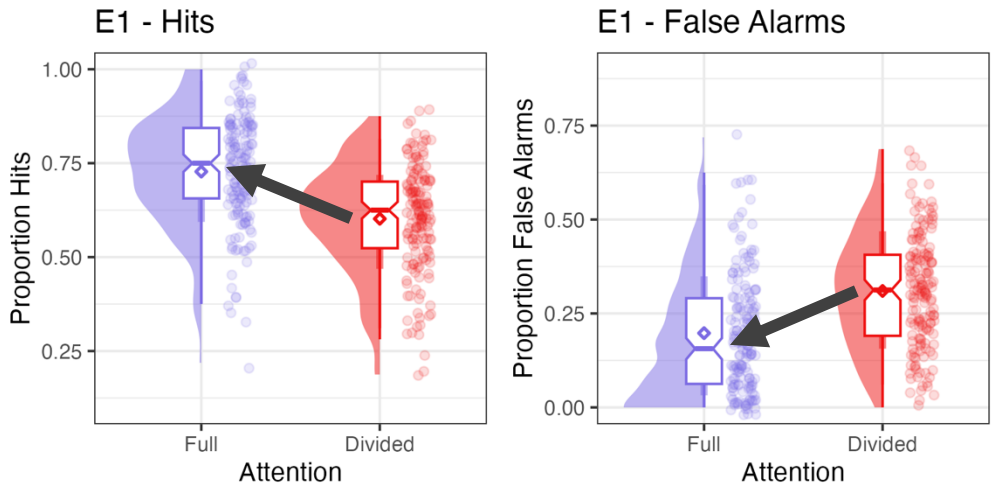
logRT: Log response time on Hits, measured from stimulus offset.

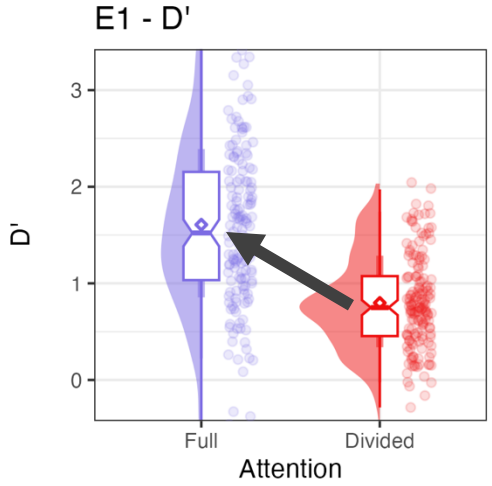


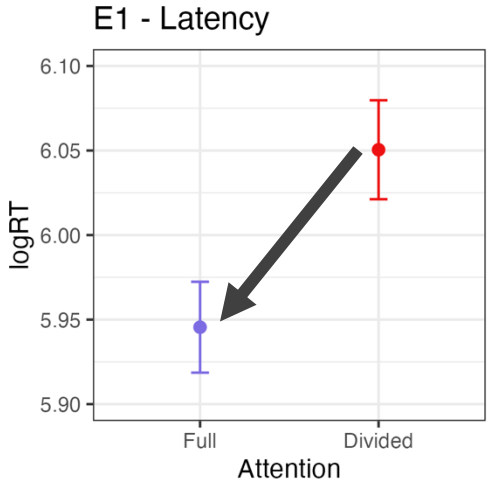
Results – Attention

More OLD sentences recognized in Full than Divided.
p < 0.001

Overall, more accurate in Full than Divided.
p < 0.001





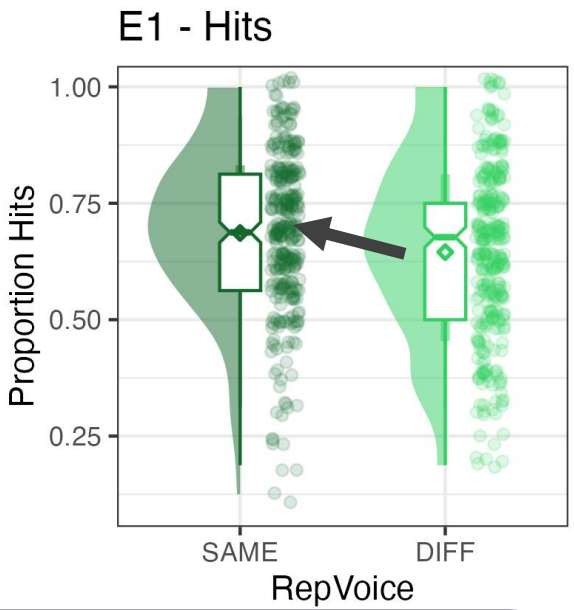


More incorrect responses on NEW trials in Divided than Full.
p < 0.001

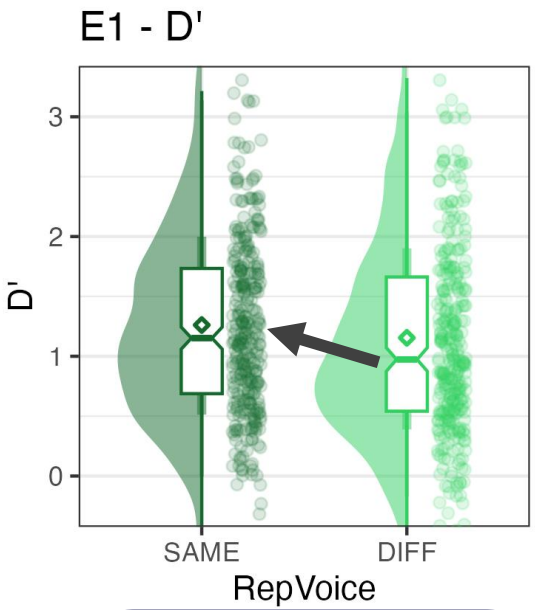
Correct responses *faster* in Full than Divided.
p < 0.05

logRT 6.1 = 445 ms
logRT 5.9 = 365 ms

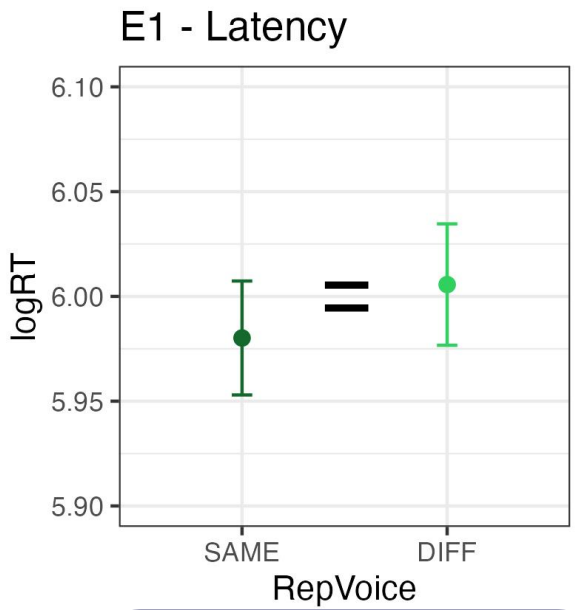
Results – RepVoice



More OLD sentences recognized when repeated by than SAME than by a DIFF talker.
p < 0.001



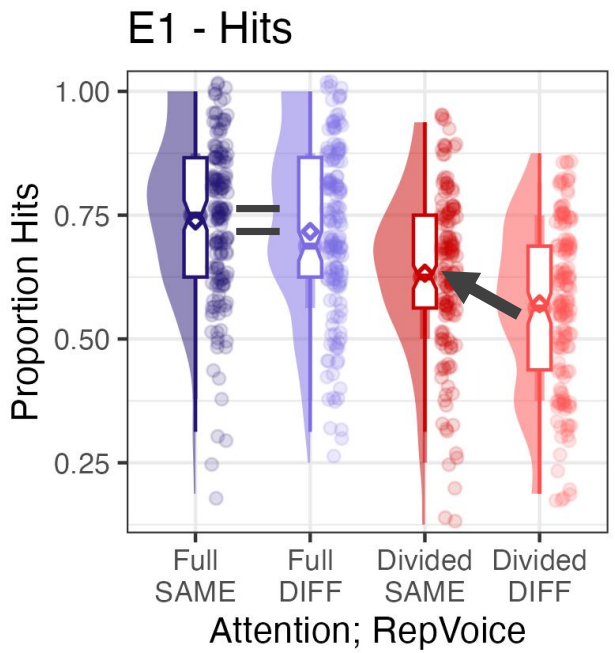
Holds after correcting for False Alarms.
p < 0.001



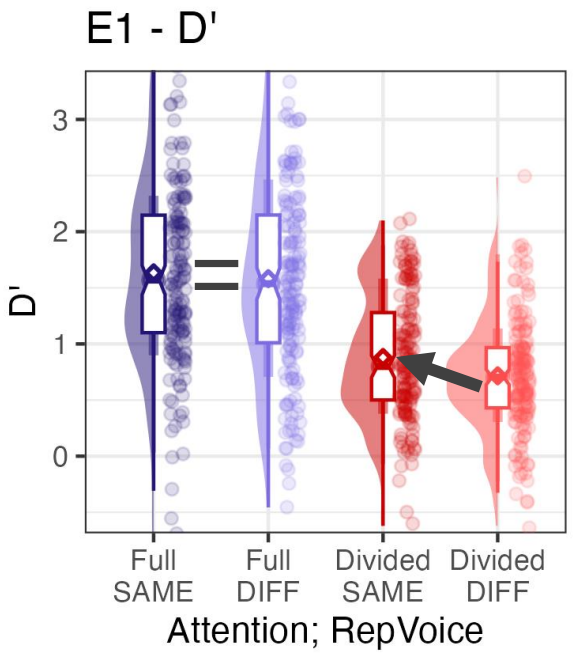
No RT advantage for SAME repetitions.
p > 0.01

logRT 6.1 = 445 ms
logRT 5.9 = 365 ms

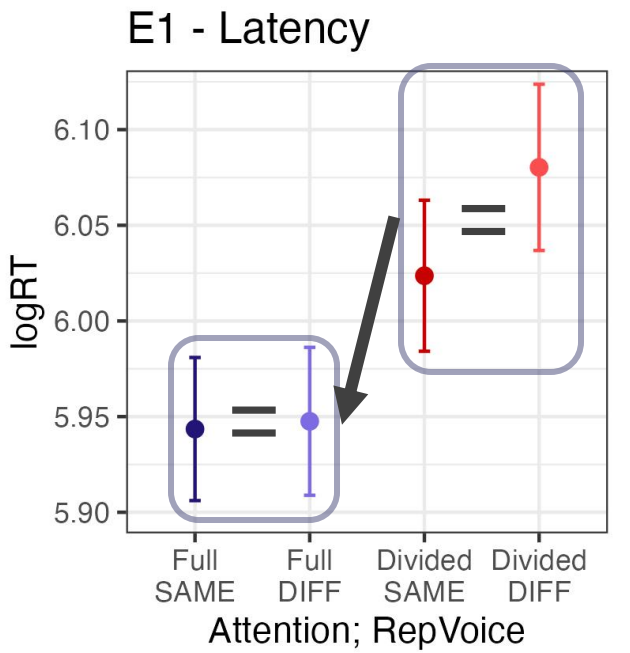
Results – Attention; RepVoice



Talker-specificity effect driven by Divided Attention. N.S. in Full.



This holds for D'.



No repVoice effect for RT.

logRT 6.1 = 445 ms
logRT 5.9 = 365 ms

How does attention at encoding affect retrieval when no acoustic cues are available?

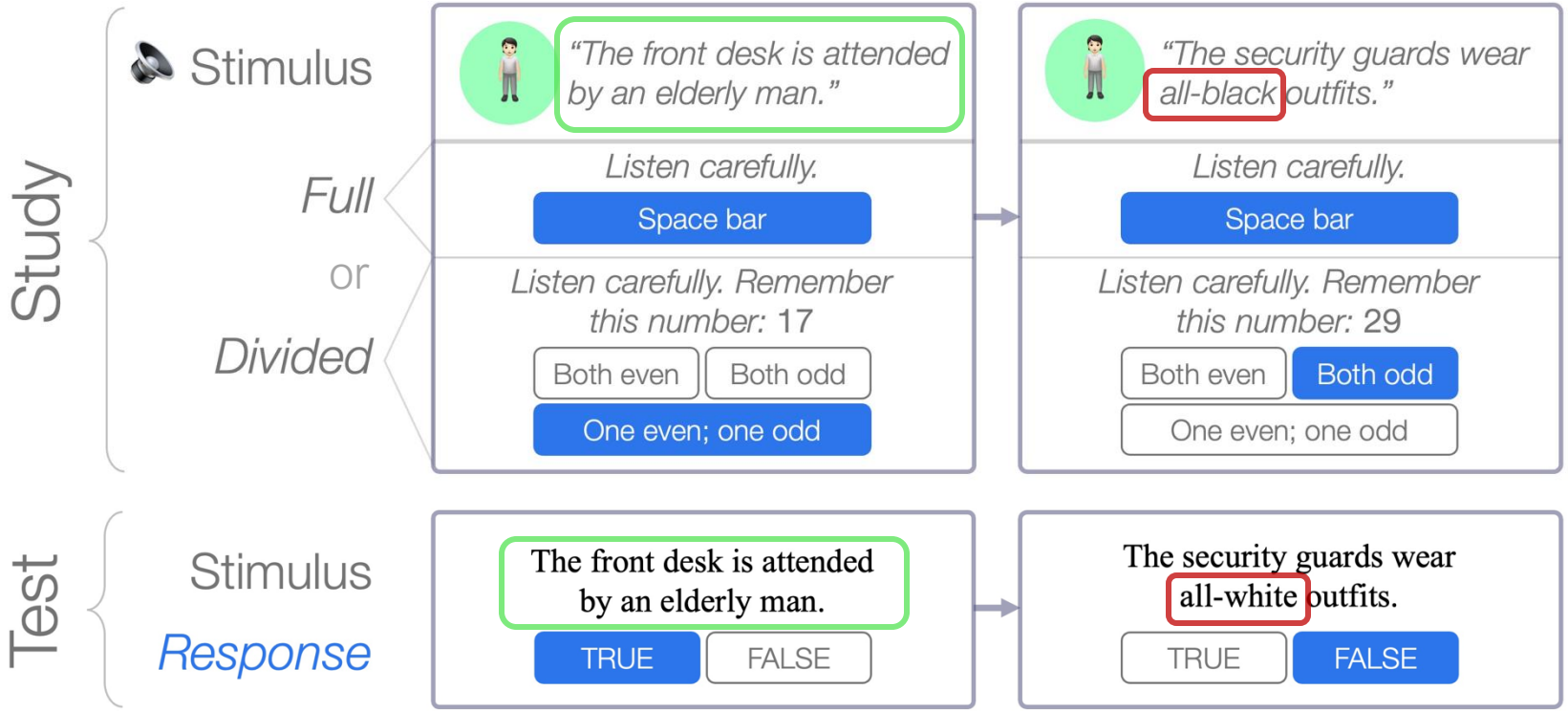
Design – Text

Full (*N* = 69)
Divided (*N* = 74)

Sentences in coherent frames: *Museum, Public Park, Train Station, Suburban Downtown.*

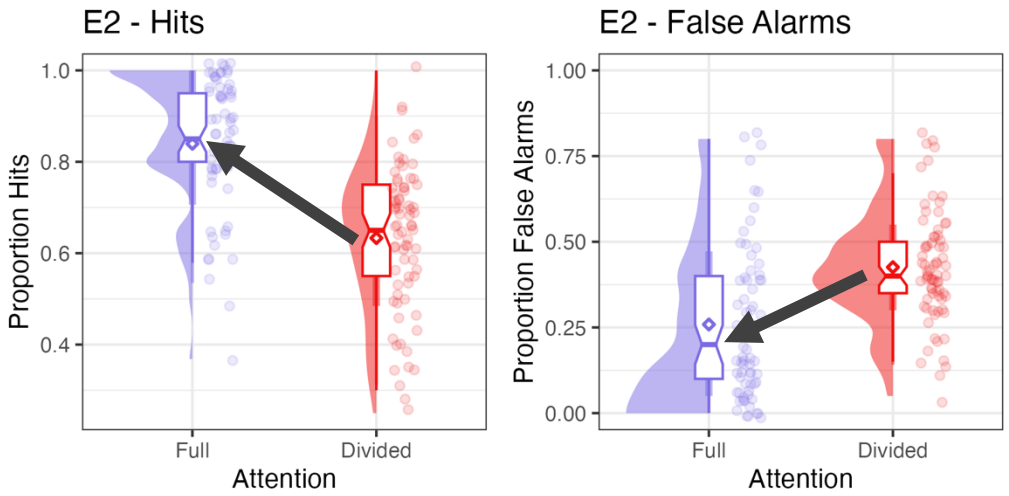
Ten semantically-opposed pairs per frame: *"The security guards wear all [black] / [white] outfits.*

Same 4 talkers. *(One per frame.)*



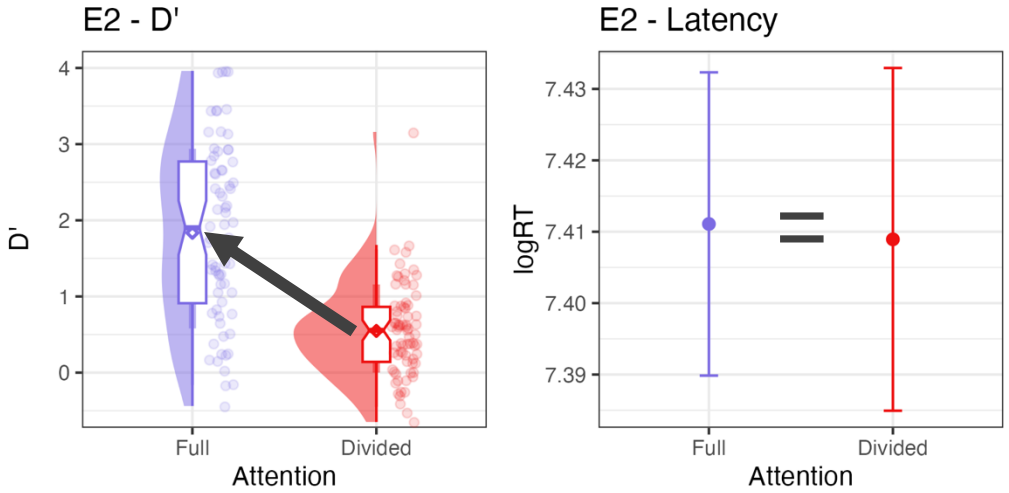
Results – Text Stimuli

More OLD sentences recognized in Full than Divided.
p < 0.001



More misidentified NEW trials in Divided than Full.
p < 0.001

Overall, more accurate in Full than Divided.
p < 0.001



No effect of attention on RTs.

logRT 7.43 = 1685 ms
logRT 7.39 = 1620 ms

How does attention at encoding influence memory for conceptual information?

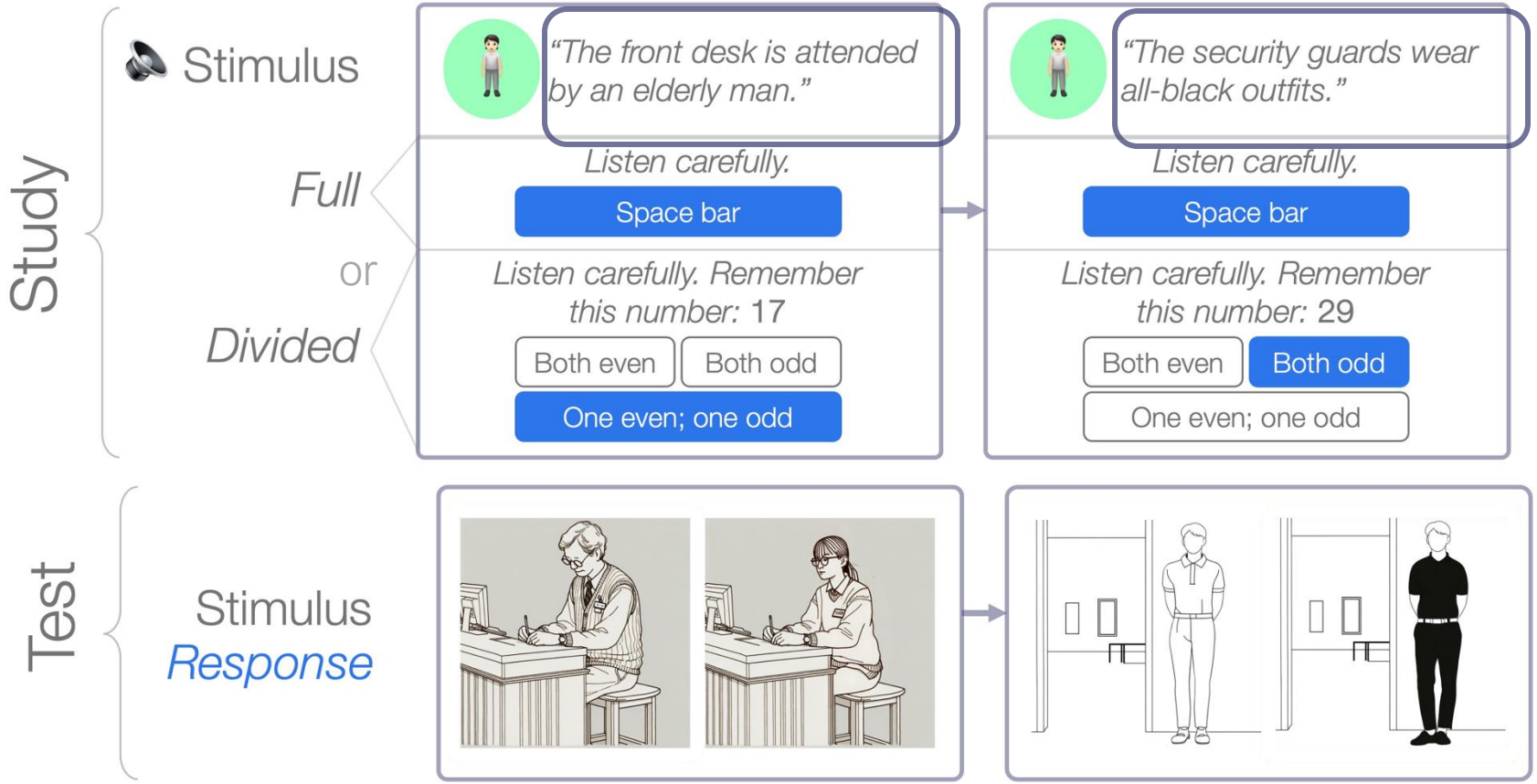
Design – Images

Full (*N* = 73)
Divided (*N* = 68)

Same study block as Exp. 2. *Same frames; same talkers.*

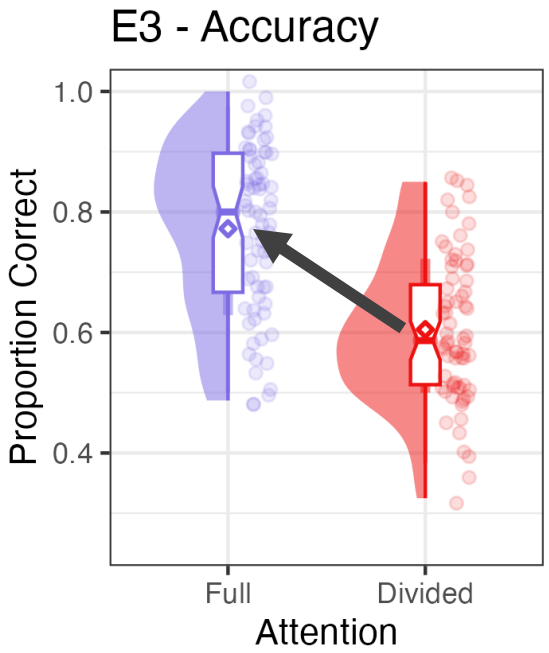
Both images presented each trial.

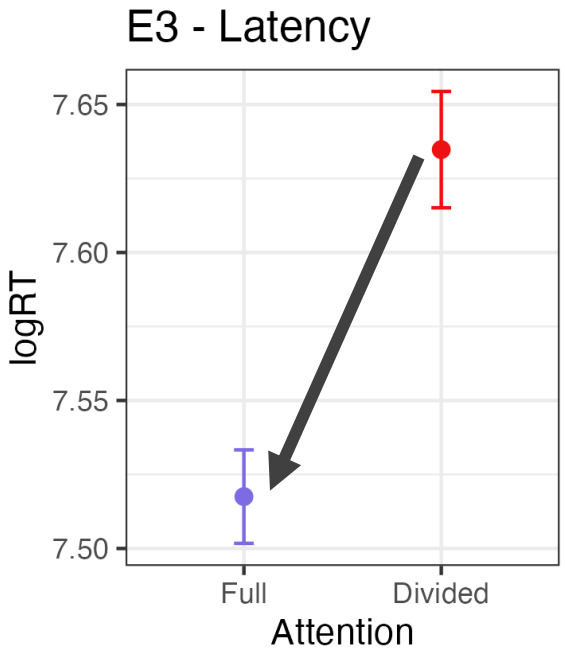
Analysis: Overall accuracy and logRT.
No traditional signal detection measures.



Results – Image Stimuli

More accurate in Full than Divided.
p < 0.001





Faster correct responses in Full than Divided.
p < 0.001

logRT 7.65 = 2100 ms
logRT 7.50 = 1808 ms

Discussion

Talker-specificity effects for ***spoken sentences.***

Effect is ***stronger*** for Divided than Full attention.

Fine-grained acoustic memory is ***fundamental*** to the system!

This info is ***not sacrificed*** when cognitive resources are scarce.

This type of encoding is ***not lexical.***

Discussion

Substantial differences across attention conditions in all Exps.

Downstream consequences of resource allocation: ***More robust*** representations of patterns we ***attentionally prioritize!***

On-the-fly resource allocation may help explain asymmetries frequency-based approaches can't account for.

*Memory for **longer utterances** is highly talker-specific!*

*Encoding of these utterances depends heavily on **resource allocation**, which likely shapes representations.*

Thank you!

Questions?
Email wsclapp@stanford.edu &
sumner@stanford.edu

Thanks to members of the Stanford Phonetics Lab and funding sources, including the NSF DDRIG, William Orr Dingwall Foundations of Language Fellowship, and Josephine de Karman Fellowship Trust.

National Science Foundation

Josephine de Karman Fellowship